Response to interactive comment on "The effect of input data complexity on the uncertainty in simulated streamflow in a humid, mountainous watershed"

Suggested title based on the reviewer comment: "Effect of input data resolution and complexity on the uncertainty of hydrological predictions in a humid, vegetated watershed"

Linh Hoang, Rajith Mukundan, Karen E. B. Moore[2], Emmet M. Owens and Tammo S. Steenhuis

Point by point response (indicated by →) to reviewers' comments

Please note that the page and line numbers referred to here are from the revised manuscript.

Anonymous Referee #2

Review of the manuscript "The effect of input data complexity on the uncertainty in simulated streamflow in a humid, mountainous watershed" by Hoang et al.

Comment

In this manuscript, Hoang et al. evaluated the effect of input resolution (digital elevation model) and input complexity (number of soil and land use classes) on model output uncertainty of the SWAT-HS model. Model output uncertainty is evaluated in terms of streamflow, saturated areas and parameter uncertainty. They conclude that uncertainty does not necessarily decrease when increasing input resolution or complexity. However, selecting parameter sets based on the combined information on streamflow and the spatial extend of saturated areas positively affected uncertainty.

This is an interesting study and I like the clear and well described concept. The results are illustrated and described in detail for different model outputs and clearly support the conclusions. The main results are well discussed. To further improve the manuscript I have some suggestions listed below. Major comments address the potential calculation of additional streamflow criteria, the calculation of a complementary measure for the saturated areas, or a figure showing simulated hydrographs. I hope that the comments below will be helpful for the authors to improve their manuscript.

Response

➔ Many thanks for the positive comment about the manuscript. Your comments are very helpful for us to improve the manuscript. We tried our best to respond to every comment and will revise the manuscript based on our responses to your comments.

1. MAJOR COMMENTS:

Comment

P L4: The model you used in this study is called SWAT-Hillslope. For me the word hillslope implies that one is working at the hillslope scale of an undisturbed catchment. However, you use the model at a much larger scale and for a catchment that is probably highly influenced by human use (urban areas and lots of agricultural area). I think it would be helpful if you shortly reflect on that and give the reader a good reason for using SWAT-HS.

Response

➔ "Hillslope" in SWAT-Hillslope does not mean hillslope scale, but means hillslope hydrology that describes paths of water through hillslope into streams. The standard SWAT does not have the ability to represent hillslope hydrology because there is no interaction between modelling units (called Hydrological Response Units, HRUs) in a subbasin. In SWAT-HS, we enabled the interaction in flow and substance transport between upland areas and the valley bottom by creating a surface aquifer. More details can be found in (Hoang et al., 2017). We will also provide a general description of SWAT-HS in supplementary materials that will be attached with the revised manuscript.

In the revised manuscript, we edited the text to clarify the meaning of hillslope in SWAT-HS as:

"SWAT-Hillslope (SWAT-HS) (Hoang et al., 2017) is a modified version of the Soil and Water Assessment Tool (SWAT) that improves the simulation of saturation-excess runoff and creates interaction in flow and substance transport between the upland areas and the valley bottom."

Comment

P5 L25-P6L8: It would be interesting to have some more information or numbers about human disturbances within the catchment: are there any major water withdrawals for agricultural use? Is there a reservoir that is used to guarantee the drinking water supply for NY in dry spells? How are these human influences affecting your model assumptions, such as a closed water balance?

Response

➔ We added some information about the Cannonsville Reservoir that the Town Brook watershed is part of the drainage area. There is no major water withdrawal for agricultural use. There is also no effort to change the hydrology in this watershed. There are a lot of activities on watershed protection programs on farms area implemented by New York City Department of Environmental Protection (NYCDEP) that helps to improve water quality, particularly phosphorus, such as cattle fencing in pastures, manure storage installation, upgrades of WWTPs. Because this manuscript only focuses on hydrology, we think it is not necessary to add these information. We will provide this information in our upcoming paper.

We revised the text in our manuscript as:

"The 37 km² Town Brook watershed is located in the Catskill Mountains, Delaware County, New York State (Fig. 1) and is the headwater of the Cannonsville Reservoir watershed which is the one of four reservoir watersheds in the New York City's Delaware system."

Comment

P6 chapter 2.2: I am not very familiar with the SWAT model and when reading the second paragraph of chapter 2.2 it was not clear to me how the structure or the combination of subbasins, wetness classes and HRUs look like. Would it be an option to include a schematic of the model structure to visually support what you are writing?

Response

➔ Thank you for the comment that reminds us about the need to provide sufficient information to non-SWAT modelers. We will provide a more detailed description of SWAT-HS setup which will include maps of input data as supplementary materials which will be attached with the revised manuscript.

Comment

P9 L2: I like the idea of using the principle of GLUE the select behavioral parameter sets. However, I am not sure if I would agree in using a Nash-Sutcliffe efficiency of 0.65 as a threshold for good simulations. How do you know that 0.65 is a good model result for your catchment? Nash-Sutcliffe is known to be high in catchments with a high discharge variability and model efficiencies also tend to be better for humid catchments than for dry catchments. Shouldn't good efficiencies for a catchment like yours be around 0.8 (I know this is a bit provocative)? Why did you decide to take a fix efficiency threshold and not just the best 10?

Response

➔ Figure 5a and figure 10a show the maximum daily NSE values in all our SWAT-HS setup which range from 0.68 to 0.69. Therefore, it is impossible to set the threshold at 0.8. From our results, daily streamflow predictions with daily NSE higher than 0.65 results in monthly streamflow prediction with monthly NSE higher than 0.8. Based on guidelines for model performance evaluation by Moriasi et al. (2007) which suggested that good model performance for streamflow corresponding to monthly NSE higher than 0.75. Therefore, we are confident that our choice of daily NSE higher than 0.65 as good model performance is a reasonable choice.

We decided to take a fixed efficiency threshold and not just the best 10 parameter sets in each setup because we would like to know how many parameter sets in each setup can give good performance above the threshold. The ratio of the number of good parameter sets of a setup to the total number of Monte Carlo parameter set (10,000 in this case) can tell us the probability of the setup to get a good model performance, which we use as one of the criteria to compare the setups.

Comment

P9 L2: Linked to using GLUE: it would be interesting to also see the simulated and observed hydrographs with the confidence intervals.

Response

➔ As the reviewer suggested, we will add the two figures below, which show the comparison between observations and simulated streamflows from all SWAT-HS setups with 90% confidence intervals, in the supplementary material that will be attached with the revised manuscript. We chose to show the comparison by plotting flow duration curves because it is very difficult to see in daily streamflow plots.

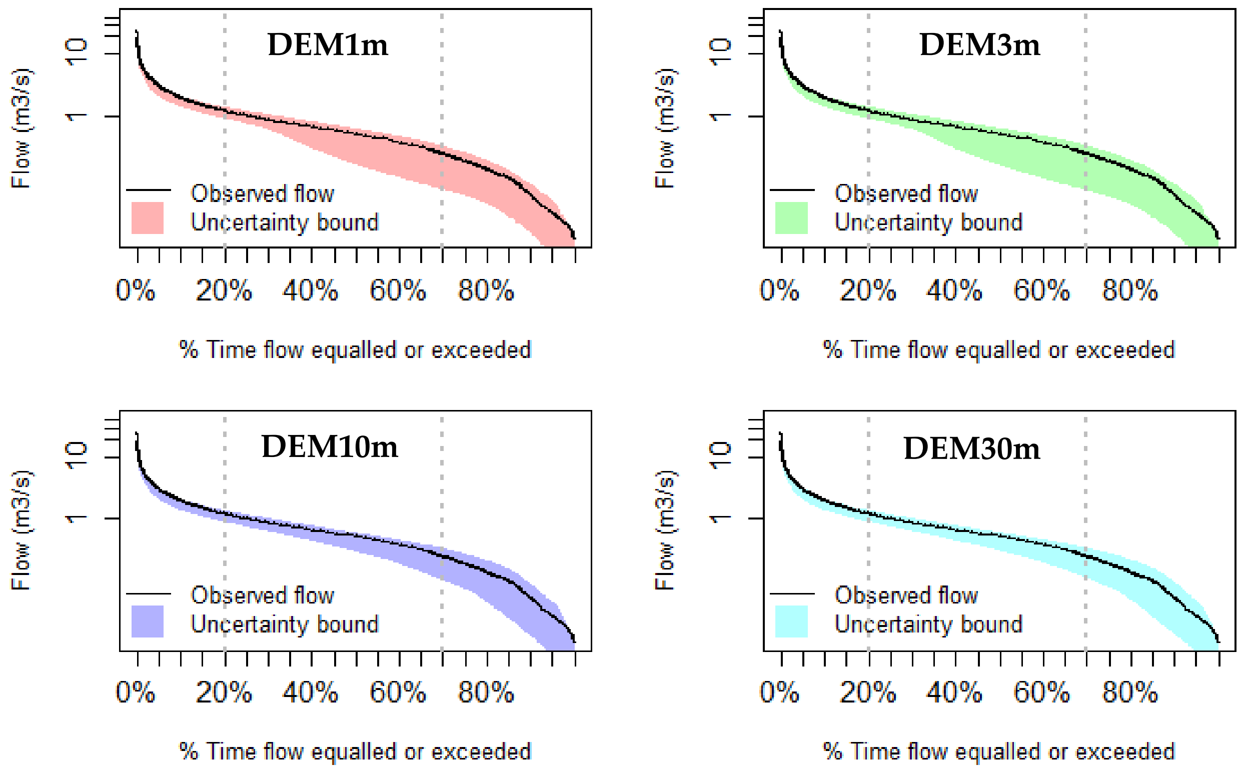

Figure 1: Uncertainty in streamflow predictions by SWAT-HS using different DEM resolutions

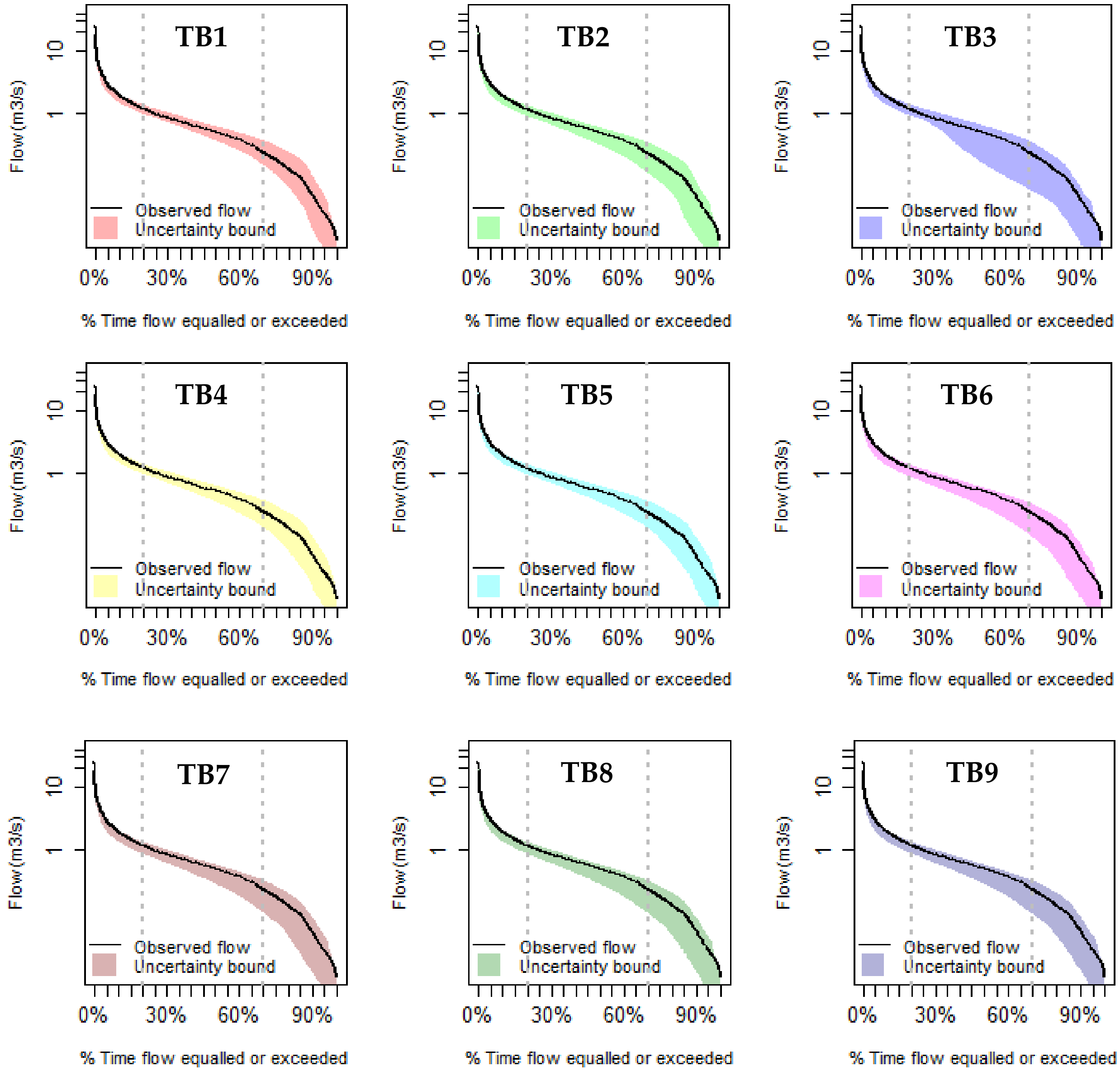

Figure 2: Uncertainty of streamflow predictions by SWAT-HS using different degrees of complexity

Comment

P9 L5: Your final goal is to make some recommendation about the appropriate model input complexity and resolution based on streamflow, saturated areas and parameter uncertainty. Evaluating the streamflow simulations on a single efficiency measure has therefore a strong impact on your final recommendations. After all the discussion on the use of Nash-Sutcliffe (e.g. Do Nash values have value from Schaefli and Gupta, 2007; Decomposition of the mean squared error and NSE performance *criteria: Implications for improving hydrological modelling from Gupta et al., 2009) I would suggest that you calculate one or two additional efficiency criteria for streamflow simulations. This could for example be a criteria representing low flow or discharge volume.*

Response

➔ Thank you very much for the reviewer's comment. We agree with your suggestion. Therefore, we calculated two additional efficiency criteria: (i) NSElog: logarithm of Nash Sutcliffe Efficiency which is a good indicator for low flows, and (ii) KGE: Kling Gupta Efficiency. We added the calculated values in table 3 and 4. We showed our edits in table 3 as an example here.

Table 3: Statistical criteria to compare the effect of DEM resolution on model uncertainty

		DEM1m	**DEM3m**	**DEM10m**	**DEM30m**
Calibration period: based on 10,000 Monte Carlo parameter sets					
Number of *"good"* parameter sets (%) for streamflow		1362	1890	2180	2293
Number of *"good"* parameter sets (%) for both streamflow and saturated areas		27	49	66	67
NSE	Max	0.69	0.69	0.69	0.69
	Mean	0.09	0.05	0.33	0.34
NSElog	Max	0.82	0.82	0.82	0.83
	Mean	0.43	0.41	0.56	0.59
KGE	Max	0.81	0.81	0.81	0.81
	Mean	0.53	0.53	0.59	0.59
Validation period: based on "good" parameter sets from calibration					
NSE	Max	0.66	0.66	0.66	0.66
	Mean	0.60	0.62	0.62	0.62
NSElog	Max	0.82	0.82	0.82	0.82
	Mean	0.70	0.70	0.69	0.71
KGE	Max	0.79	0.78	0.79	0.79
	Mean	0.70	0.70	0.70	0.71

Comment

P14 L1: If I understand correctly, you evaluate the model simulations by the percentage of simulated areas that intersect with the observed areas (purple color in Fig. 8). This corresponds to my interpretation to the percentage of correct classifications. To me it seems logical that the DEM30m performs best, because it cannot be too wrong due to its coarse resolution. Therefore, I think that evaluating the percentage area of misclassification (percentage of simulated area that does not intersect with the observation green color in Fig. 8) would give additional and important information for the evaluation of the various DEM resolutions.

Response

➔ We confirm that the reviewer understood correctly that we evaluated the model simulation by comparing the percentage of simulated areas that intersects with the observed areas. We also think that it is logical that DEM30m has the highest percentage due to its coarse resolution. However, the most important reason is that the coarse resolution DEMs (DEM 10m and DEM 30m) gave a realistic distribution of topographic index (TI) values with the high TI grids well compatible to the stream network. Thus, with the classification of wetness classes based on TI values, coarse resolution DEMs also provide a better distribution of wetness classes with the highest TI wetness classes (which supposed to be 'wet') locating in the downslope, near-stream areas while the lowest TI wetness classes ('dry' wetness classes) being in the upslope areas.

In our opinion, we would like to focus on the percentage of simulated areas that intersects with the observed areas (the correct classification) rather than the percentage that does not intersect with observation (the misclassification) to evaluate the model simulation. We think that the correct classification reflects the model performance which we are evaluating. The misclassification is easily to calculate by deducting the correct classification from 100%.

We will provide a proper explanation the reason why coarser resolution gave a better prediction of saturated areas in the revised version.

Comment

P18 L2: In the discussion you provide some good reasons why a relatively coarse DEM resolution (DEM 10) can lead to good/ acceptable results. However, I don't understand why a better resolution does not result in even better results. Could you maybe write a few sentences that elaborate on that?

Response

➔ Same as our response to your previous comment, a relatively coarse DEM resolution (DEM 10m) can lead to a better performance it gives a more realistic distribution of topographic index (TI) values which results in a better distribution of wetness classes. The reason is explained as below.

Note that the basic equation for topographic index is TI = ln (contributing area/slope angle)

The below figure shows the relationships of TI with slope angle, upslope contributing area and elevation using 2 representative DEM resolutions: 1m and 10m. It is clearly observed that DEM 1m can capture a significantly wider range of slope than DEM 10m because of its finer resolution. Also, the percentage of grids that has low values of TI is significantly higher in DEM 1m than in DEM 10m (in figure below use red lines for reference), which also can be seen in figure 3d in the main manuscript. Low TI values are usually found in grids with steep slope or with low upslope contributing areas. Because DEM 1m captures steep slope at local scale and has a high number of grids with low upslope contributing area (figure 3c in the main manuscript), the percentage of low TI values in DEM 1m is much higher. If we look at the relationship between TI and elevation, we can see that the distribution of TI values in DEM 1m spread out wider than in DEM10m at all elevations. This explains why the distribution of wetness classes in DEM1m has a more complex pattern with every wetness class spread-out while DEM10m has a more coherent pattern with high TI wetness class well compatible with the stream network (Figure 4 in the main manuscript).

Our findings are in agreement with Lane et al. (2004) who used high resolution LiDAR 2m DEM in the TOPMODEL which simulates hydrology based on topographic index. The TOPMODEL predicted the widespread existence of disconnected saturated zones that expand within an individual storm event but which do not necessarily connect with the drainage network. They found that using the LiDAR 2m DEM, the topographic index has a complex pattern, associated with small areas of both low and high values of the topographic index, leading to the appearance of disconnected saturated areas. After remapping the topographic data are remapped at progressively coarser resolutions by spatial averaging of elevations within each cell, they found that as the topographic resolution is coarsened, the number and extent of unconnected saturated areas are reduced: the catchments display more coherent patterns, with saturated areas more effectively connected to the channel network. Moreover, in another study, Quinn et al. (1995) showed how progressively fining model resolution from 50 m to 5 m reduces the kurtosis in the distribution of topographic index values and increases quite substantially the number of very low index values.

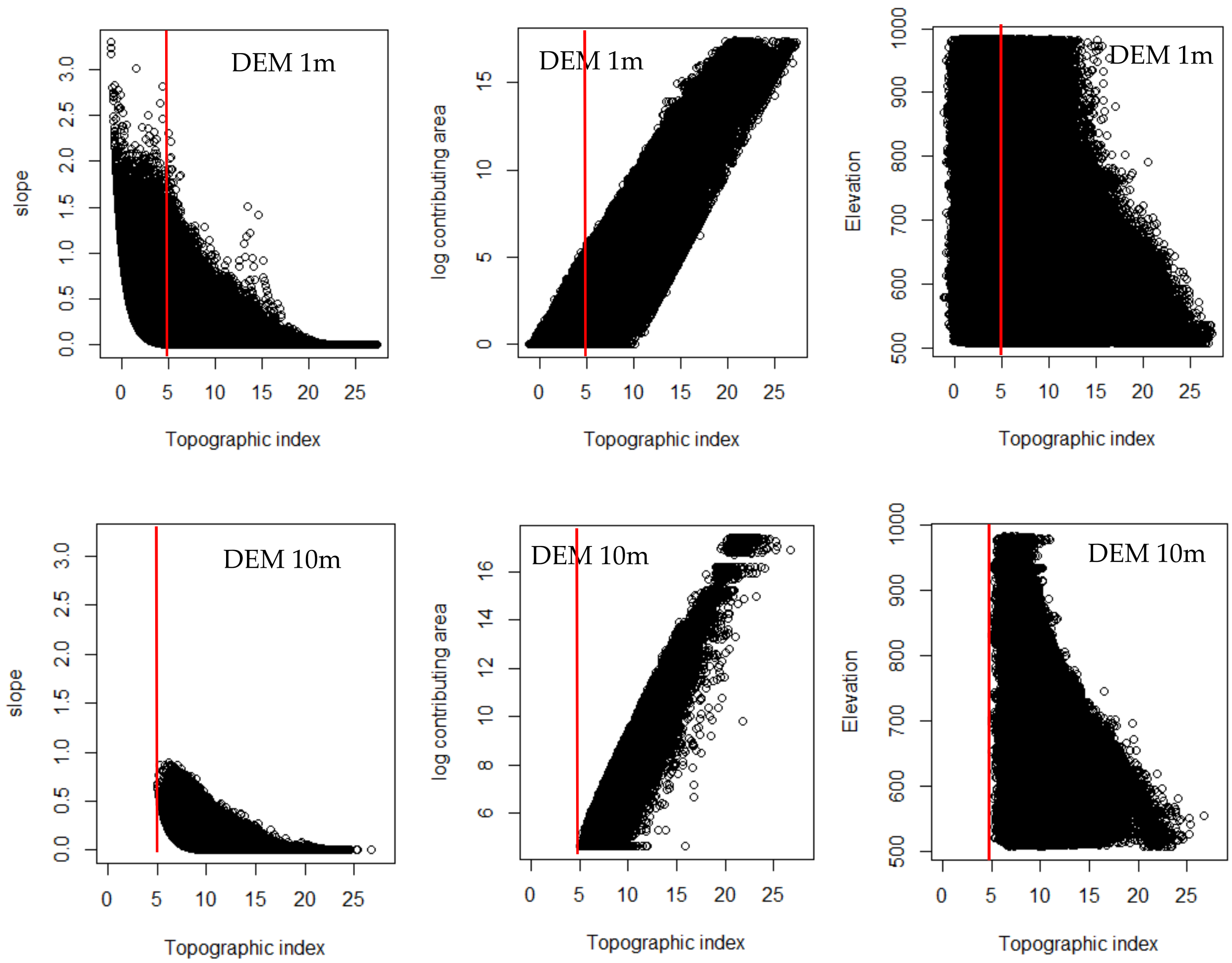

Figure 3: Relationship of topographic index with slope, upslope contributing area and elevation with two resolution of DEM: 1m and 10m

Comment

Table 3 and 4: I agree that the number of parameter sets and the median efficiency values are interesting. However, I would recommend to add the values directly to the corresponding figures to have the information where it is relevant. I think that max and min efficiency values can also be seen/ guessed from the figures and are not that important that they need to be in a table.

Response

➔ Based on your previous suggesting to add more efficiency criteria, we added the values of maximum and mean NSElog and KGE in these two tables. We removed median NSE because the readers can read those values from figure 5 and 10.

2. MINOR COMMENTS:

Comment

P1 L1: After reading the manuscript I would suggest to adapt the title, because is only addresses part of the actual analysis. I am also not sure if the study catchment can be considered as mountainous. So maybe the title could be adapted to something similar as: Effect of input data resolution and complexity on simulation uncertainty for a simple runoff model.

Response

➔ Thank you for the comment. We will revise our title as: "**Effect of input data resolution and complexity on the uncertainty of hydrological predictions in a humid, vegetated watershed**

Comment

P1 L11: I would be careful with using the term "water quality" in the very first sentence of the abstract as it suggests that the study is about water quality, which is not the case.

Response

➔ We agree with the reviewer. In the revised manuscript, we removed the term "water quality" as:

"Uncertainty in hydrological modelling is of significant concern due to its effects on prediction and subsequent application in watershed management."

Comment

P2 L4: The nine model setups not only had a similar effect on parameter uncertainty, but also on streamflow simulation.

Response

➔ In this sentence, we mentioned about the effect of input data complexity first while the following sentence is about the effect on streamflow simulation. Therefore, we did not make any edit regarding this comment in the revised manuscript.

Comment

P2 L7: The term spatial input details was a bit confusing to me. I would rather use the same terms as before: input resolution and complexity.

Response

➔ We changed the term "increasing spatial input details" to "improving input resolution and complexity".

Comment

P2 L16-17: Is data used to calibrate the model (e.g. discharge data) included in your list? For me it would be a fourth point called output data uncertainty.

Response

➔ In this study, we do not consider measurement error as uncertainty. Model uncertainty from the three components (model structure, input data and model parameters) that we mentioned are caused by assumptions or simplifications made for model structure, averaging impact or lack of data in input data and the effect of equifinality on model parameters. We assumed that all available measurements for input data and data used for calibration are true. Therefore, measured data uncertainty is not in the list. Moreover, this categorization of model uncertainty is not our own but based on Lindenschmidt et al. (2007).

Comment

P5 L12-22: I recommend to reorganize this paragraph. Having two listings in a row makes it more difficult to understand what the focus of your study is.

Response

➔ We reorganized the paragraph and hope the focus of the study is clear now.

The text was edited in the revised version as:

"The two main objectives of this paper are to evaluate: (i) the effect of DEMs of various spatial resolution (1, 3, 10, and 30 m) on the uncertainty of streamflow and saturated area predictions, and (ii) the impact of combinations of soil and land use data with various degrees of complexity on the uncertainty in model simulation. In both analyses, we not only investigate the effect on model prediction/output uncertainty but also discuss their effect on the uncertainty in parameter estimation. Through this study we seek to answer specific questions including identifying the suitable DEM resolution in order to get good model performance, and the appropriate complexity of the distributed input data. Answers to these research questions will be the basis for reducing decision uncertainty on model input selection in our future applications of SWAT-HS in the NYC water supply system."

Comment

P7 L4: I recommend to mention the concept of "hydrological connectivity", which is the argument for you lateral surface aquifer.

Response

➔ We agree with the reviewer. We will provide a more detailed description of SWAT-HS which includes the explanation of the '"hydrological connectivity" concept in more detail as supplementary material attached to the revised manuscript.

Comment

P7 L20: Please add the reference for the LiDAR data.

Response

➔ We added the reference for the LiDAR data in the revised manuscript as:

"The 1m DEM (DEM1m) was derived from 2009 aerial LiDAR data acquired by New York City Department of Environmental Protection (RACNE, 2011)."

Comment

P7 L25. I would refer to Figure 4. (": : :divided into 10 wetness classes (Fig. 4))

Response

➔ We edited the text as:

"Based on TI values, the watershed was divided into 10 wetness classes (Figure 4)."

Comment

P8 L 9: Please add the reference for the solar radiation data.

Response

➔ We added the reference for the solar radiation data and revised the text as:

"Solar radiation data was derived as the average of airport stations at Albany and Binghamton as supplied from the Northeast Regional Climate Center."

Comment

P8 L21: Using a model with a snow routine I would mention the percentage of precipitation falling as snow in the section where you describe the catchment.

Response

➔ We added the information of percentage of precipitation falling as snow in the case study description as:

"Approximately 1/3 of the total precipitation in the region fall as snow (Pradhanang et al., 2011).

Comment

P8 L19: Why do make 10'000 MC runs and not 100'000? Do you think your parameter distributions would look differently with more random parameter sets? Please give some reasons for your choice.

Response

➔ The calibration procedure in this manuscript is based on our previous study (Hoang et al., 2017). Our calibration procedure includes 2 stages: snowmelt calibration (5 parameters) and flow calibration (9 parameters among which 1 parameter, ALPHA_BF, is not very sensitive based on our previous study). In each stage, we generated 10,000 parameter sets which were run with SWAT-HS and the results were compared with observations.

We assume that we generate 1 random sample in at least high, middle, and low range for each parameter. In the stage of snow melt calibration, with 5 parameters involved, 3^5 = 243 is the minimum number of MC parameter sets required to cover the parameter space. In the flow calibration stage, with 8 sensitive parameters involved, 6561 combinations are the minimum number required. Therefore, we think that using 10,000 MC parameter sets for each stage of calibration is sufficient.

Comment

P10 chapter 3.1.1: You could think about moving this chapter to the methods part.

Response

➔ We would like to keep this section in the *Results* part because this is a part of our analysis which is the basis to explain the effect of DEMs on prediction of saturated areas. In the *Methodology* part in the revised manuscript, we mentioned about this work as:

"We evaluated the effect of DEM resolution on representing topographical characteristics of the watershed by comparing the statistical distributions of elevation, slope angle, upslope contributing area, and TI using DEMs with various spatial resolutions (1m, 3m, 10m and 30m)."

Comment

P13 L7: Could you briefly explain what the percentage of saturated areas is? It would then also become clearer what / how many data points the corresponding boxplots (Fig. 7) contain.

Response

➔ The percentage of saturated areas is defined as the percentage of the watershed area that is saturated in the simulated day. The number of data points for each DEM in figure 7 equals to the number of days in the calibration period (2556 days) multiplying the number of good parameters for both streamflow and saturated areas in each DEM setup.

Comment

Fig. 1: Could you increase the resolution of this figure? Because it is not sharp when printing it out on A4.

Response

We provided a high resolution figure for figure 1 in the revised manuscript (see Figure 1).

Comment

Fig. 5 and 10: I recommend to adapt the y axis scales to better use the available space. And I also suggest to use the same style/ content of figure caption for Fig. 5 and Fig. 10.

Response

➔ Thanks the reviewer for the detailed comment.

We edited the caption for Figure 5a and 5b and the general caption of figure 5 to be consistent with figure 10 (see figure 5). Moreover, we also adjusted the y axis scales of plots in figure 5 and 10 as the reviewer's suggestion (see Figure 5 and Figure 10).

Comment

Fig. 6: Again, I recommend to adapt the y axis scales to better use the available space. Additionally, I would add the information that only the good parameter sets for both streamflow and saturated areas are used in this plot to the figure caption.

Response

➔ We would like to keep this figure as it is. The reason is that this figure does not only show variations of probability of saturation in each wetness class using different DEMs, it also aims at comparing the difference of these variations between wetness classes. Therefore, we purposely kept similar value ranges for y axis in all plots.

As the reviewer suggested, we changed the caption of figure 6 to: *Probability of saturation of wetness classes in SWAT-HS set ups with different DEM resolutions using good parameters for both streamflow and saturated areas*

Comment

Fig. 9 and 11: It is almost impossible to see the distributions of the good parameter sets for both streamflow and saturated areas. Why don't you scale the y axis differently?

Response

➔ We agree with the reviewer. We changed the y axis of figure 9 and 11 to logarithmic scale. The new figure 9 is shown below as an example of our edits.

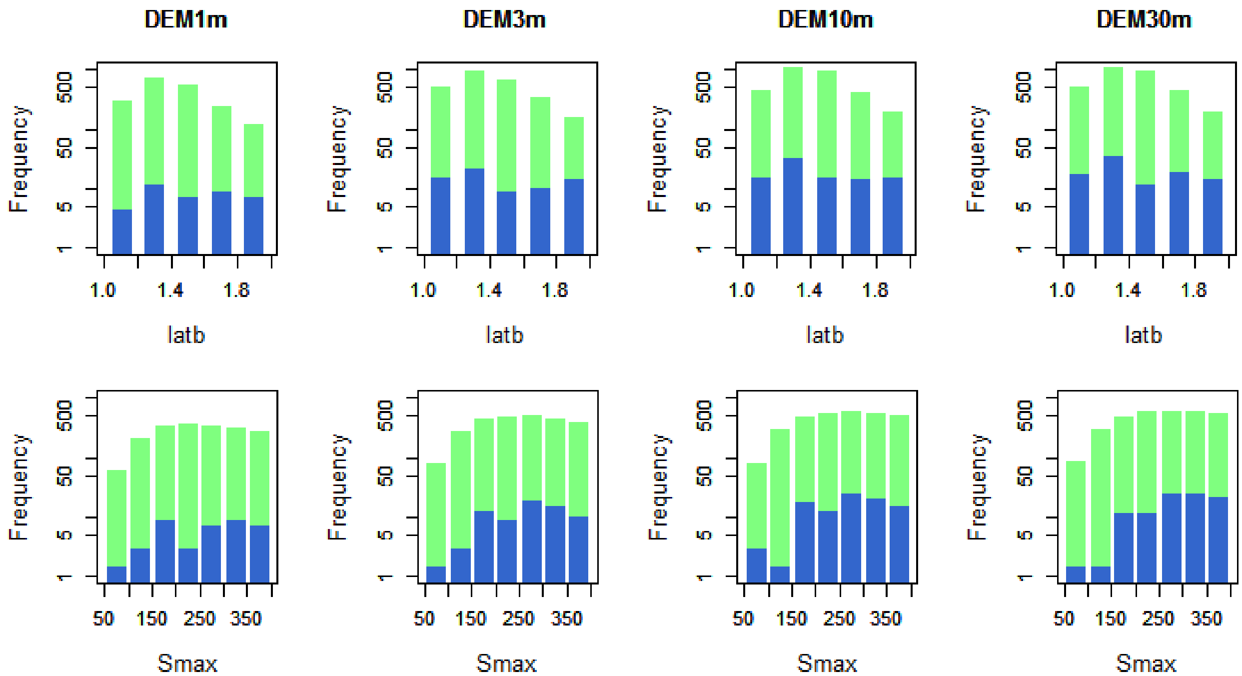

Figure 9: Distribution of *"good"* parameters for streamflow (in green) and for both streamflow and saturated areas (in blue) with log y axis in four SWAT-HS setups using different DEM resolutions

Comment

Table 1: What is the difference between latA and latB? It is not clear since the definitions are identical.

Response

➔ The lateral flow in SWAT-HS is generated as the following equation using the linear (*latA*) and exponential (*latB)* coefficients :

$$\overline{latQ} = latA * \overline{S_1} \wedge latB$$

where ($\overline{latQ}$) is lateral flow for the sub-basin, $\overline{S_1}$ is the amount of water stored in the surface aquifer, *latA* and *latB* are constant coefficients.

The equation can be found in Hoang et al. (2017). We also added a description of SWAT-HS in the supplementary material attached to the revised manuscript.

Comment

Generally: I suggest to use the HESS guidelines for making references and also for formatting units.

Response

➔ We edited our manuscript to follow the HESS requirement of units and reference style.

REFERENCES

Hoang, L., Schneiderman, E. M., Moore, K. E. B., Mukundan, R., Owens, E. M., and Steenhuis, T. S.: Predicting saturation-excess runoff distribution with a lumped hillslope model: SWAT-HS, Hydrological Processes, 31, 2226-2243, 10.1002/hyp.11179, 2017.

Lane, S. N., Brookes, C. J., Kirkby, M. J., and Holden, J.: A network-index-based version of TOPMODEL for use with high-resolution digital topographic data, Hydrological Processes, 18, 191-201, doi:10.1002/hyp.5208, 2004.

Lindenschmidt, K.-E., Fleischbein, K., and Baborowski, M.: Structural uncertainty in a river water quality modelling system, Ecological Modelling, 204, 289-300, 2007.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., and Veith, T. L.: Model evaluation guidelines for systematic quantification of accuracy in watershed simulations, Transactions of the ASABE, 50, 885-900, 2007.

Pradhanang, S. M., Anandhi, A., Mukundan, R., Zion, M. S., Pierson, D. C., Schneiderman, E. M., Matonse, A., and Frei, A.: Application of SWAT model to assess snowpack development and streamflow in the Cannonsville watershed, New York, USA, Hydrological Processes, 25, 3268-3277, doi:10.1002/hyp.8171, 2011.

Quinn, P. F., Beven, K. J., and Lamb, R.: The in(a/tan/β) index: How to calculate it and how to use it within the topmodel framework, Hydrological Processes, 9, 161-182, doi:10.1002/hyp.3360090204, 1995.

RACNE: CAT-393: Airborne LiDAR GIS Terrain and Hydrology Data Development - (Phase 2), Draft report, 2011.